Replacement Abstract:

A novel extension of the vector space model for computing chemical similarity is described. In one embodiment, a method calculates similarity between molecules and molecular descriptors using the singular value composition (SVD) of a molecule/descriptor matrix and, for example, an identity matrix, to create a low dimensional representation of the original descriptor space. Probe or query molecules then can be projected into the low dimensional representation and compared to the molecules from the original matrix.